TITLE OF THE INVENTION

IMAGE PROCESSING APPARATUS, IMAGE PROCESSING METHOD, IMAGE
PROCESSING PROGRAM, AND COMPUTER-READABLE STORAGE MEDIUM
STORING IMAGE PROCESSING PROGRAM CODE

BACKGROUND OF THE INVENTION

Field of the Invention

[0001]   The present invention relates to an image
processing apparatus, an image processing method, an image
processing program, and a computer-readable storage medium
storing image processing program code, and more particularly,
to image-data sending/receiving processing according to the
traffic on a communication network.

Description of the Related Art

[0002]   Cellular telephones (or portable terminals) are
now being widely used. Fig. 1 illustrates a typical example
of a communication system using portable terminals.

[0003]   In Fig. 1, portable terminals 401 and 405 each
include a display unit, an operation unit, and a
communication controller, and communicate with a relay
device (base station) 403 via communication channels 402 and
404.

[0004]   Modulation methods are rapidly shifting towards

the use of digital data rather than analog data, and
portable terminals are being increasingly used not only as
telephones for sending and receiving audio data, but also as
portable terminals for sending and receiving data.

5   Additionally, along with increases in the transmission rate,
video (moving pictures) can be sent and received, which is
impossible in the known art, and it is expected that such
portable terminals will be used as video phones.

[0005]   Fig. 2 is a block diagram illustrating a known
10 video phone system. In Fig. 2, a video camera 501 captures
an image of, for example, a character, and outputs a video
signal. A microphone 504 receives sound and outputs an
audio signal.

[0006]   Analog-to-digital (A/D) converters 502 and 505
15 convert the signals output from the video camera 501 and the
microphone 504, respectively, into digital signals.

[0007]   A video encoder 503 encodes the digital video
signal output from the A/D converter 502 according to a
known compression/encoding method, and an audio encoder 506

20 encodes the digital audio signal output from the A/D
converter 505 according to a known compression/encoding
method. Generally, compressed and encoded data is referred
to as a "bitstream".

[0008]   A multiplexer 507 multiplexes the video bitstream
25 and the audio bitstream so that they can be played back in

synchronization with each other, thereby generating a single bitstream.

[0009]   A demultiplexer 508 demultiplexes the multiplexed bitstream into the video bitstream and the audio bitstream. Then, a video decoder 509 decodes the video bitstream, a digital-to-analog (D/A) converter 510 converts the digital video data into an analog signal, and a monitor 511 displays the decoded video signal.

[0010]   An audio decoder 512 decodes the audio bitstream, a D/A converter 513 converts the digital audio data into an analog signal, and a speaker 514 outputs the decoded sound.

[0011]   A communication controller 515 sends and receives the above-described bitstreams.   Reference numeral 516 indicates a communication channel, which is a wireless channel in this example.   A relay device (base station) 517 sends and receives data to and from portable terminals. Reference numeral 518 indicates a communication channel 518 via which the relay device 517 and the portable terminals communicate with each other.   A synchronization controller 519 controls the video signal and the audio signal to be played back in synchronization with each other by using timing control information superposed on each bitstream.

[0012]   In the above-described known video phone system, however, depending on the traffic on the communication network, pictures or sound may not be continuously received

at a receiving side, thereby failing to properly transmit information.

SUMMARY OF THE INVENTION

[0013]   Accordingly, in view of the above background, the present invention has been made in order to solve the above-described problem. It is an object of the present invention to provide an image processing apparatus and method for implementing data communication such that images can be transmitted and received continuously regardless of the traffic status on a communication network.

[0014]   In order to achieve the above object, according to one aspect of the present invention, there is provided an image processing apparatus including: a natural-image input unit for inputting a natural-image signal obtained by encoding a natural image; an artificial-image input unit for inputting an artificial-image signal obtained by encoding an artificial image; and a transmitter for adaptively multiplexing the natural-image signal and the artificial-image signal according to a communication status of a communication network, and for transmitting a resulting multiplexed signal via the communication network.

[0015]   According to another aspect of the present invention, there is provided an image processing apparatus

for decoding a multiplexed signal obtained by adaptively multiplexing an encoded natural-image signal and an encoded artificial-image signal according to a communication status of a communication network. The image processing apparatus includes: a receiver for receiving the multiplexed signal; a separator for separating the received multiplexed signal into the natural-image signal and the artificial-image signal; a natural-image decoder for decoding the natural-image signal separated by the separator; and an artificial-image decoder for decoding the artificial-image signal separated by the separator.

[0016] According to still another aspect of the present invention, there is provided an image processing method including: a natural-image input step of inputting a natural-image signal obtained by encoding a natural image; an artificial-image input step of inputting an artificial-image signal obtained by encoding an artificial image; and a transmission step of adaptively multiplexing the natural-image signal and the artificial-image signal according to a communication status of a communication network, and transmitting a resulting multiplexed signal via the communication network.

[0017] According to a further aspect of the present invention, there is provided an image processing method for decoding a multiplexed signal obtained by adaptively

multiplexing an encoded natural-image signal and an encoded artificial-image signal according to a communication status of a communication network. The image processing method includes: a receiving step of receiving the multiplexed signal; a separation step of separating the received multiplexed signal into the natural-image signal and the artificial-image signal; a natural-image decoding step of decoding the separated natural-image signal; and an artificial-image decoding step of decoding the separated artificial-image signal.

[0018]   According to a yet further aspect of the present invention, there are provided a computer-readable storage medium in which computer-readable program code implementing the above-described image processing method is stored and program software for controlling a computer to execute the above-described image processing method.

[0019]   Other objects, features, and advantages of the invention will become apparent from the following detailed description taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0020]   Fig. 1 illustrates an example of a communication system using portable terminals.

[0021]   Fig. 2 is a block diagram illustrating the configuration of a known video phone system.

[0022]   Fig. 3 is a block diagram illustrating the configuration of a video phone system according to a first embodiment of the present invention.

[0023]   Figs. 4A through 4D illustrate examples of synthesized images.

[0024]   Fig. 5 is a block diagram illustrating a detailed configuration of a multiplexer 110 shown in Fig. 3.

[0025]   Fig. 6 illustrates a mesh representing a graphics skeleton.

[0026]   Fig. 7 illustrates an example of a face image model.

[0027]   Fig. 8 is a block diagram illustrating the configuration of a video phone system according to a second embodiment of the present invention.

[0028]   Fig. 9 is a block diagram illustrating the configuration of a video phone system according to a third embodiment of the present invention.

[0029]   Fig. 10 illustrates the total bit rate when video data and animation data are synthesized so as to form the synthesized images shown in Figs. 4A through 4D.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0030]   The present invention is described in detail below with reference to the accompanying drawings through illustration of preferred embodiments.

First Embodiment

[0031]   Fig. 3 is a block diagram illustrating the configuration of a video phone system according to a first embodiment of the present invention.

[0032]   The video phone system includes a transmitter and a receiver. In Fig. 3, in the transmitter, a video camera 101 for capturing natural images and outputting video data (natural-image video data), an A/D converter 102, a video encoder 103, a microphone 104, an A/D converter 105, and an audio encoder 106 are similar to the video camera 501, the A/D converter 502, the video encoder 503, the microphone 504, the A/D converter 505, and the audio encoder 506, respectively, shown in Fig. 2, and an explanation thereof is thus omitted. The video encoder 103 performs encoding processing in compliance with, for example, the ISO/IEC 14496-2 (MPEG-4 Visual) standards.

[0033]   A communication controller 115, a communication channel 116, a relay device 117, and a communication channel 118 are also similar to the communication controller 515, the communication channel 516, the relay device 517, and the communication channel 518, respectively, shown in Fig. 2,

and an explanation thereof is thus omitted.

[0034]   An animation generator 119 in the transmitter generates animation data (artificial-image data) in response to instructions from an operation unit 130. The animation generator 119 generates graphical animation data (including skeleton data, movement data, and texture data, which will be discussed below) obtained by simulating, for example, the expression of a face and the movement of hands. A technique for creating animation is described below.

[0035]   An animation encoder 120 compresses and encodes the animation data generated by the animation generator 119.

[0036]   A multiplexer 107 adaptively selects the output of the video encoder 103 (video stream) and the output of the animation encoder 120 (animation stream) in response to instructions from the operation unit 130, and multiplexes the video stream and the animation stream, thereby outputting an image stream.

[0037]   A multiplexer 121 multiplexes the image stream output from the multiplexer 107 and an audio stream output from the audio encoder 106, and supplies a multiplexed data steam to the communication controller 115.

[0038]   In the receiver, a demultiplexer 122 demultiplexes the data stream input from the communication controller 115 into an image stream consisting of the video data and/or the animation data and an audio stream based on attribute

information stored in the header of the data stream.

[0039]   A demultiplexer 108 demultiplexes the video data and the animation data from the image stream based on attribute information stored in the header of the image stream.

[0040]   The video data, the animation data, and the audio data are decoded by a video decoder 109, an animation decoder 123, and an audio decoder 112, respectively. A D/A converter 113 converts the audio data decoded by the audio decoder 112 into analog data. A speaker 114 outputs the analog audio data.

[0041]   An animation synthesizer 124 processes the animation data decoded by the animation decoder 123 by synthesizing, for example, the face and the hands of a character image. A synchronization controller 111 controls the video data or the animation data to be synchronized with the audio data.

[0042]   A multiplexer 110 determines how the video data and/or the animation data have been multiplexed and transmitted in the transmitter, and synthesizes the video data and the animation data based on the result of the determination, thereby outputting synthesized image data to a display controller 125. Details of the multiplexer 110 are given below. The video data and/or the animation data are displayed on a monitor 126.

[0043]   In the first embodiment, the type of image to be synthesized from video data (natural image) and animation data (artificial image) can be selected from a plurality of types by the operation unit 130. Examples of the types of synthesized images are shown in Figs. 4A through 4D.

[0044]   Fig. 4A illustrates a synthesized image using only video data (natural image) output from the video camera 101 for both the background image and the character image. Fig. 4B illustrates a synthesized image using animation data (artificial image) generated by the animation generator 119 for the background image and using video data output from the video camera 101 for the character image. Fig. 4C is a synthesized image using video data output from the video camera 101 for the background image and using animation data generated by the animation generator 119 for the character image. Fig. 4D is a synthesized image using only animation data generated by the animation generator 119 for both the background image and the character image.

[0045]   The synthesizing processing performed by the multiplexer 110 is described below with reference to Fig. 5.

[0046]   Video data output from the video decoder 109 is temporarily stored in a primary frame buffer 1000. Normally, video data is two-dimensional pixel data handled in units of frames. In contrast, animation data using polygons is usually three-dimensional data. Thus, video data and

animation data cannot be synthesized without further
processing.

[0047]   Accordingly, after animation data is generated in
the animation synthesizer 124, it is temporarily stored in a
two-dimensional primary frame buffer 1001, and then,
rendering is performed on the animation data, thereby
constructing frame data.

[0048]   If the animation data is used for the background
image (see Fig. 4B), it is combined with the video data used
for the foreground image by using mask information of the
video data (which is obtained by a masking information
controller 1003) in units of frames. If the animation data
is used for the foreground image (see Fig. 4C), a mask image
is formed from a two-dimensional video image by performing
rendering, and then, the animation data is combined with the
video data based on the mask image.

[0049]   The animation synthesizing speed is synchronized
with the video playback speed in a frame controller 1002.
The frame data formed in the primary frame buffers 1000 and
1001 and the mask information obtained by the masking
information controller 1003 are input into a frame
synthesizer 1004. Then, two frames (or a greater number of
primary frames) are combined while suitably performing
masking processing by using the mask information, and the
resulting synthesized image is written into a display frame

buffer 1005. As a result, a natural image consisting of video data and animation data can be output.

[0050] The technique for creating animation data used in this embodiment is as follows. Fig. 6 illustrates a mesh indicating a graphics skeleton.

[0051] As stated above, the graphics skeleton is referred to as a "mesh". In the mesh shown in Fig. 6, each unit formed by connecting vertices (in the shape of a triangle in the example shown in Fig. 6) is generally referred to as a "polygon". For example, the portion formed by connecting vertices A, B, and C is defined as one polygon.

[0052] In order to construct the graphics shown in Fig. 6, it is necessary to indicate the coordinate values of the individual vertices and information concerning combinations of vertices (for example, vertices A, B, and C, vertices A, G, and H, and vertices A, E, and D). Although the above type of graphics is generally constructed in a three-dimensional space, the ISO/IEC 14496-1 (MPEG-4 Systems) defines the above type of configuration in a two-dimensional space.

[0053] In practice, image (pattern) data referred to as a "texture" is mapped onto each polygon of the graphics skeleton. This is referred to as "texture mapping". Then, a realistic-looking graphical model can be formed.

[0054] Motion can be added to the graphics object shown

in Fig. 6 by changing the coordinate positions of the polygons over time, as indicated by the arrows shown in Fig. 6. If the directions and the magnitudes of the individual vertices are the same, a simple translation operation is implemented. By changing the magnitude and the direction for the individual vertices, motion and the transformation of the graphics object is possible.

[0055] If motion information concerning the individual vertices is constantly re-defined, the amount of data becomes large. Accordingly, only a difference in motion vectors of each vertex is recorded, or the translation time and the translation locus are pre-defined, and an object is automatically animated along the locus in an animation device according to predetermined rules.

[0056] The animation creating technique is more specifically discussed below in the context of a face image. Fig. 7 illustrates an example of a face image model.

[0057] Unlike a general graphics object, a face model has common features concerning the eyes, the nose, etc. The model shown in Fig. 7 is formed of parameters consisting of A: the distance between the center of one eye to the center of the other eye, B: the vertical length of the eyes, C: the vertical length of the nose, D: the length from the bottom line of the nose to the top line of the mouth, and E: the horizontal length of the mouth.

[0058]   By preparing a plurality of sets of parameters and a plurality of corresponding textures, a template set for face animation can be formed. As stated above, in a face image, there are many common feature points representing, for example, the corners of the eyes, and the corners of the mouth. By changing the positions of the feature points, many facial expressions can be created.

[0059]   For example, by providing commands, such as "lower the position of the feature points representing the corners of the eyes" (in practice, configuration data near the feature points are also changed), and "lift the positions representing the corners of the mouth", a "smiling" expression can be created.

[0060]   Accordingly, the number of bits per unit time required for transmitting graphics data is smaller than that for transmitting moving pictures.

[0061]   The above-described animation creating technique is also applicable to a body image, as well as a face image. More specifically, feature-point data representing, for example, joints for the hands and the feet, is extracted, and motion information is added to the extracted data, thereby making it possible to animate actions, such as "walking" or "lifting a hand", with a small amount of data.

[0062]   According to the first embodiment, a data stream obtained by suitably combining video data and animation data

in one screen can be sent and received in response to user instructions. Thus, the bit rate of the data stream can be controlled by changing the ratio of the video data to the animation data to be synthesized. With this arrangement,

5 the data stream can be sent and received according to the traffic status.

Second Embodiment

[0063]   Fig. 8 is a block diagram illustrating the

10 configuration of a video phone system according to a second embodiment of the present invention. In Fig. 8, the elements having the same functions as those shown in Fig. 3 are designated with like reference numerals, and an explanation thereof is thus omitted.

15 [0064]   In Fig. 8, an animation-template storage unit 201 stores template information (skeleton, complexion, hair-style, with or without glasses) for face animation data. An animation selector 202 selects the animation template and the animation pattern (such as "waving a hand" or "lowering

20 the head") according to the user's taste.

[0065]   That is, in the second embodiment, a plurality of templates for animation data are prepared, and the user suitably selects a desired template so as to create animation data and transmit it.

25 [0066]   According to the second embodiment, the user is

able to easily create animation having a desired motion, and a data stream obtained by suitably combining video data and the created animation data in one screen can be sent and received in response to user instructions. Thus, the bit

5 rate of the data stream can be controlled by changing the ratio of the video data to the animation data. As a result, the data stream can be sent and received according to the traffic status.

Third Embodiment

10 [0067] Fig. 9 is a block diagram illustrating the configuration of a video phone system according to a third embodiment of the present invention. In Fig. 9, the same elements having the same functions as those shown in Fig. 8

15 are indicated by like reference numerals, and an explanation thereof is thus omitted.

[0068] In Fig. 9, a video tracker 301 is a device for identifying and extracting a certain object (for example, a face) from video data by using a suitable method.

20 [0069] A video analyzer 302 analyzes the object image extracted by the video tracker 301 so as to analyze the individual objects forming the video data, and supplies the analysis result to an animation selector 202'.

[0070] For example, in analyzing a human being, the video

25 analyzer 202' analyzes the contour of the face, the

positions of the eyeballs, the position of the mouth, etc.

[0071]   A communication status monitor 303 monitors the communication status (such as the effective bit rate and the traffic congestion) of the communication channel, and controls animation to be generated, the animation data to be adaptively multiplexed with the video data, and the synthesized bitstream to be transmitted according to the communication status.

[0072]   The synthesizing processing of video data and animation data according to the communication status is discussed below with reference to Figs. 4A through 4D. In Figs. 4A through 4D, it is assumed that the foreground image (character) is moving rapidly and the background image is stationary. Fig. 10 illustrates the total bit rate when video data and animation data are combined so as to form the synthesized images shown in Figs. 4A through 4D. In Fig. 10, (a), (b), (c), and (d) respectively correspond to the images shown in Figs. 4A, 4B, 4C, and 4D.

[0073]   In this embodiment, When the communication status is good (for example, higher bit-rate data can be sent and received since a data channel is unoccupied), only video images are used (Fig. 4A or (a) of Fig. 10), and as the communication status becomes worse (for example, only lower bit-rate data can be sent and received since a data channel becomes congested), the synthesizing processing is

adaptively and automatically controlled to form a
synthesized image according to the ratio of the video data
to the animation data in the order of Fig. 4B, 4C, and 4D,
or (b), (c), and (d) of Fig. 10.

[0074]   The animation selector 202' selects the animation
template according to the result obtained from the
communication status monitor 303 and the video analyzer 302,
thereby generating realistic-looking animation.

[0075]   As described above, video data and animation data
are combined to form one screen with a suitable ratio (see
Figs. 4A through 4D) according to the communication status
(the ratio of the video data to the animation data changes
according to the communication status). Additionally, sound
can also be transmitted according to the user's taste.

[0076]   According to the third embodiment, video data and
animation data can be adaptively multiplexed and transmitted
according to the communication status, thereby preventing
interruptions of images or sound at the receiving side.

[0077]   The bit rate of the animation data itself can also
be reduced by dynamically decreasing the resolution of a
mesh forming the animation data. By using this technique,
the bit rate can be further reduced according to the traffic
status.

[0078]   Software program code for implementing the
functions of the foregoing embodiments may be supplied, and

may be executed by a program stored in a computer (or a CPU or an MPU). The present invention also encompasses such a modification.

[0079]   In this case, the program code itself implements the novel functions of the foregoing embodiments. Accordingly, the program code itself, and means for supplying such program code to the computer, for example, a storage medium storing such program code, constitute the present invention. Examples of the storage medium for storing and supplying the program code include a floppy disk, a hard disk, an optical disc, a magneto-optical disk, a compact disc read only memory (CD-ROM), a CD-recordable (CD-R), a magnetic tape, a non-volatile memory card, and a read only memory (ROM).

[0080]   In other words, the foregoing description of the embodiments has been given for illustrative purposes only and is not to be construed as imposing any kind of limitation.

[0081]   The scope of the invention is, therefore, to be determined solely by the following claims and not limited by the text of the specification and alterations made within a scope equivalent to the scope of the claims fall within the true spirit and scope of the invention.